基于叶绿体基因组 SNP 的天台鹅耳枥谱系结构与分化分析

陈模舜，杨仲毅

（台州学院 生命科学学院 浙江省植物进化生态学与保护重点实验室，浙江 台州 318000）

摘要：天台鹅耳枥为中国特有的濒危植物，仅间断分布于浙江省境内，种群数量稀少，已处于极危状态。该文通过对 6 个自然居群（包含所有居群的母株）cpDNA 单核苷酸多态性（SNP）研究，探讨天台鹅耳枥谱系结构与系统分化，以评估濒危状况，并提出相应的保护策略。使用 TIANGEN 试剂盒法提取基因组 DNA，用 Illumina novaseq 6000 进行高通量测序，对获得叶绿体全基因组序列，使用在线程序 OGDRAW 制作叶绿体基因组图谱，用 DnaSP 分析核苷酸多样性，用 PopART 软件进行单倍型网络构建，使用 RAxML 软件构建 ML tree，用 MrBayes 构建 Bayes tree。结果表明：（1）通过天台鹅耳枥叶绿体全基因组序列分析，大多数蛋白质编码基因和氨基酸序列显示出明显的密码子偏好，检测到 cp LTR 正向重复为 32 个，回文重复为 25 个，反转重复为 22 个；SSR 重复序列不同类型为 87 个，其中大多数富含 A / T，单核苷酸的数量最多。（2）在叶绿体基因组中鉴定了 314 个单核苷酸多态性，单核苷酸取代显示天台鹅耳枥群体属单系，分为天台县居群（THS）和景宁县居群（JST），居群单倍型之间演化关系呈现星状中心辐射。（3）所有居群核苷酸多样性的变异均较低（P_i＜0.005），JST 居群和 THS 居群单倍型多样性较低（H_d 为 0.5～0.6），显示出天台鹅耳枥在历史上遇到瓶颈后曾发生局部扩张，居群间呈现较大的遗传分化，居群内具有较低的遗传变异与居群间较高的分化水平。通过对叶绿体基因组 SNP 的研究，揭示天台鹅耳枥的遗传多样性和谱系分化，为濒危植物天台鹅耳枥种质资源保护和遗传拯救提供理论依据。

关键词：天台鹅耳枥，叶绿体全基因组，单核苷酸多态性，谱系结构，系统发育

中图分类号：Q943；S 431.18　　　　　　　　**文献标识码：**A

Genealogical structure and differentiation analysis of *Carpinus tientaiensis* based on single nucleotide polymorphism of chloroplast genome

CHEN Moshun, YANG Zhongyi

基金项目：浙江省基础公益研究计划项目（LY19C060001）[Supported by Basic Public Welfare Research Project of Zhejiang Province (LY19C060001)]。

第一作者：陈模舜（1962-），学士，副教授，主要从植物生态学、种群遗传学、濒危植物保护生物学研究，（E-mail）cmshoh@tzc.edu.cn。

(*School of Life Sciences*, *Taizhou University*, *Zhejiang Provincial Key Laboratory of Plant Evolutionary Ecology and Conservation*, Taizhou 318000, Zhejiang, China)

Abstract: *Carpinus tientaiensis* is an endangered plant species with sparse population densities and is endemic to China. The distribution of this species is restricted to the territory of Zhejiang Province and its survival rate is very low. Based on data of chloroplast (cp) single nucleotide polymorphism (SNP) in six natural populations (including the mother plants of all populations), genealogical structure and systematic development, we assessed how endangered *C. tientaiensis* is and put forward relevant conservation strategies. The genomic DNA of *C. tientaiensis* was extracted using TIANGEN kit method and a library was constructed for high-throughput sequencing with Illumina novaseq 6000. To obtain the complete sequence of chloroplast genome, chloroplast genome maps were constructed using the online program OGDRAW. Nucleotide diversity was analyzed by DnaSP software while PopART was used for haploid network construction and RAxML was adopted for ML tree construction. Finally, MrBayes software was used for construction of Bayes tree. The data showed that: (1) According to whole-genome sequence analysis of *C. tientaiensis*, most of the protein-coding genes and amino acid sequences showed obvious codon preference. Moreover, long-term relationship (LTR) revealed 32 forward repeats, 25 palindrome repeats and 22 reverse repeats. We detected 87 Simple Sequence Repeats (SSR) in different types, most of which were A/T rich and the number of single nucleotides was the largest. (2) A total of 314 SNPs was identified in the cpDNA of *C. tientaiensis*. Single-nucleotide substitution indicated that *C. tientaiensis* was monophyletic and could be categorized into Tiantai county (THS) and Jingning county (JST) populations. The evolution of the haplotypes of the population manifested stellar radiation. (3) The variation of nucleotide diversity was low in all populations ($P_i < 0.005$) and the haplotype diversity of THS and JST populations was also low (H_d 0.5-0.6). This indicated that *C. tientaiensis* had undergone historically local expansions after experiencing bottlenecks across its evolution history, with lower genetic variation within populations and higher level of diversity among populations. The results of SNP of the chloroplast genome revealed the genetic diversity and lineage differentiation of *C. tientaiensis*, and provided theoretical basis for conservation of the genetic resources as well as rescue of this endangered plant species.

Keywords: *Carpinus tientaiensis*, chloroplast genome, single nucleotide polymorphism, genealogical structure, phylogeny

天台鹅耳枥（*Carpinus tientaiensis*）属于桦木科（Betulaceae）鹅耳枥属（*Carpinus*），为中国特有和濒危的第三纪孑遗植物，是国家二级重点保护野生植物（章绍尧等，1993）。我国是鹅耳枥属植物的分布中心，约有 33 种、8 变种（傅立国，2003），其中东南沿海是鹅耳枥属植物一个重要的分布区，天台鹅耳枥（*C. tientaiensis*）和普陀鹅耳枥（*C. putoensis*）等为浙江地区特有（陈之端，1994）。中国植物志曾记载天台鹅耳枥产于浙江东部（中国科学院中国植物志编辑委员会，1979），浙江植物志统计天台鹅耳枥仅存 5 株（章绍尧等，1993），但近年来在磐安县和景宁县等地有所发现，仅间断分布于浙江省境内的天台县、磐安县、青田县和景宁畲族自治县，野外成年植株不足 50 株，幼年个体缺少，低于野外种群稳定存活界限，已处于极危状态。生境片断化使得野生居群变小及居群间的隔离程度增加，这将会导致遗传变异丧失和近交衰退，最终增加物种灭绝的风险（Aguilar et al.，

2008； Wei et al.，2012）。根据 IUCN 物种红色名录，天台鹅耳枥属于极危（CR）等级。天台鹅耳枥染色体为 14 倍体（2n=14x=112），是桦木科倍性最高的多倍体植物（陈模舜等，2020）。天台鹅耳枥在研究桦木科分类、古植物区系、濒危机制等研究中具有很高的科研价值（章绍尧等，1993；王昌腾等, 2007）。

目前仅对天台鹅耳枥组织解剖结构、光合特性对生长光强的响应、群落特征等方面进行了研究（陈模舜等，2013，2020）。但天台鹅耳枥种群的进化机制与种内系统发育尚未明确，必须收集群体样本进行基因组学的分析，在基因组水平上加深对这一濒危物种居群动态的了解。叶绿体基因组由于其高基因含量和保守的基因组结构，可用于研究开花植物的母系遗传，特别是多倍体植物（Birky, 1995；Soltis et al., 2000）。叶绿体基因组通常包含单亲遗传的 DNA，由于其自我复制机制和相对独立的进化，来自叶绿体的遗传信息经常被用来探索近缘种间和种内的亲缘关系（于涛等，2019）。目前叶绿体全基因组高通量测序技术为系统进化分析提供了大量的信息位点，比较基因组结构、基因的变异和重复序列排列，有利于构建居群遗传结构、居群历史动态和谱系间分化（孙逸，2012；王佳慧，2015）。本研究通过对天台鹅耳枥叶绿体基因组高通量测序，通过对单核苷酸多态性（SNP）研究，分析多态位点和核苷酸变异，评估居群的遗传多样性水平，推断天台鹅耳枥的谱系结构和分化，为天台鹅耳枥种质资源的保护与恢复制定策略。

1 材料与方法

1.1 天台鹅耳枥样地调查及测试材料

天台鹅耳枥自然居群最高海拔相近，均在 890 m 以上，同属亚热带山地湿润气候，土壤 pH 值为酸性，森林植被覆盖率较高，植被类型为亚热带常绿阔叶林（陈模舜等，2020）。天台鹅耳枥分布范围狭窄，仅浙江省的天台县华顶山、磐安县的大盘山、高姥山和青明尖，青田县仰天湖和景宁畲族自治县上山头存在野生群落。其中天台县华顶山共存 19 株，青田县仰天湖仅存 1 株，景宁县大际乡上山头胸径 20 cm 以上 18 株，磐安县境内的大盘山共存 3 株，高姥山和青明尖各为 1 株。磐安县每个居群之间相隔 20 km 之内，天台县居群与景宁县居群直线距离约 250 km（表 1）。

表 1 天台鹅耳枥分布地的基本概况

Table 1 Basic conditions of each spot of *Carpinus tientaiensis*

采集编号 Collecting number	样本量 Sample size	分布地区 Distribution region	经纬度 Latitude and longitude	海拔高度 Altitude (m)	样本编号 Sample ID
JST	10	景宁县上山头 Shangshantou, Jingning County	119°37′ E， 27°46′ N	1 520	JST_1～10
PDS	3	磐安县大盘山 Dapan Mountain, Pan’an County	120°31′ E，28°58′ N	1 130	PDS_1～3

PGS	1	磐安县高姥山 Gaomu Mountain, Pan'an County	120°32′ E 28°54′ N	1 140	PGS_1
PQJ	1	磐安县青明尖 Qingmingjian, Pan'an County	120°28′ E，28°49′ N	1 200	PQJ_1
QYH	1	青田县仰天湖 Yangtianhu, Qingtian County	119°59′ E，28°12′ N	1 246	QYH_1
THS	10	天台县华顶山 Huading Mountain, Tiantai County	121°05′ E，29°15′ N	890～1 000	THS_1～10

天台鹅耳枥材料采集自 6 个自然居群 26 个植株（包含所有居群的母株）（表 1）。其中景宁县上山头样地 50 m ×15 m 每隔 5～6 m 采集 1 株，共 9 株平均胸径 25.2 cm；孤树 1 株，分枝最大胸径 11.1 cm。磐安县大盘山、高姥山、青明尖全部采集共 5 株，平均胸径 20.41 cm。青田县仰天湖采集 1 株，胸径 28.66 cm。天台县华顶山西茅棚采集 7 株平均胸径 43.08 cm，公路附近 3 株平均胸径 20.07 cm。野外采集新鲜叶子后将其冷冻在干冰中，清洁叶片，并保存在-80 ℃的冰箱中以备后续实验研究。

1.2 叶绿体基因组测序、组装与注释

叶绿体基因组 DNA 提取使用 TIANGEN 的 DNAsecure 新型植物基因组 DNA 提取试剂盒（DP320），然后在 Illumina novaseq 6000 平台，以 150 bp 的对端读数高通量测序，每个样本至少 6 Gb 的原始序列数据。在过滤原始数据并消除数据质量的影响（Phred 分数 Cutoff-30）之后，我们获得了高质量的数据。对获得 26 个天台鹅耳枥叶绿体全基因组序列，利用已发表的叶绿体基因组序列（登录号：KY174338）作为参考序列（Yang et al., 2017），采用 DOGMA 软件注释完整的叶绿体基因组（Wyman et al., 2004），使用在线程序 OGDRAW（https://chlorobox.mpimp-golm.mpg.de/OGDraw.html）制作叶绿体基因组图谱，同时对每条序列的 cp 基因组基本信息进行统计，包括基因组大小、基因特征和 GC 含量。

1.3 重复序列、IR 区域与边界的扩张收缩和密码子偏好性分析

重复序列中散在重复使用 Reputer 软件分析，用 cpLTR 表示；简单重复序列（SSR）鉴定使用 MISA（http://pgrc.ipk-gatersleben.de/misa/）工具（参数：1-10 2-5 3-4 4-3 5-3 6-3）分析。比较不同序列 LSC、SSC 和 IR 长度及边界，使用在线工具 IRscope（https://irscope.shinyapps.io/irapp/）完成 IR 区域与边界的扩张收缩的分析。通过 R 软件分析密码子偏好并作图。

1.4 系统发育树分析

为了检查物种不同区域的系统发育，根据完整的 cp DNA 序列进行系统发育分析，使用 RAxML8.0 软件进行极大似然树（ML tree）构建，结合 MrBayes 3.3 软件进行贝叶斯推断法构建 Bayes tree。基于叶绿体基因组开发单倍型，运用 PopART 软件构建单倍型网络（Kimura, 1980；Leigh et al., 2015）。使用 DnaSP v6（http://www.ub.edu/dnasp/）软件分析核苷酸多样性参数，通过 AMOVA 分子方差分析，评估谱系间的分子变异程度和遗传分化

指数 F_{st}，分析得到谱系间的分化程度。

2 结果与分析

2.1 天台鹅耳枥叶绿体的基本特征

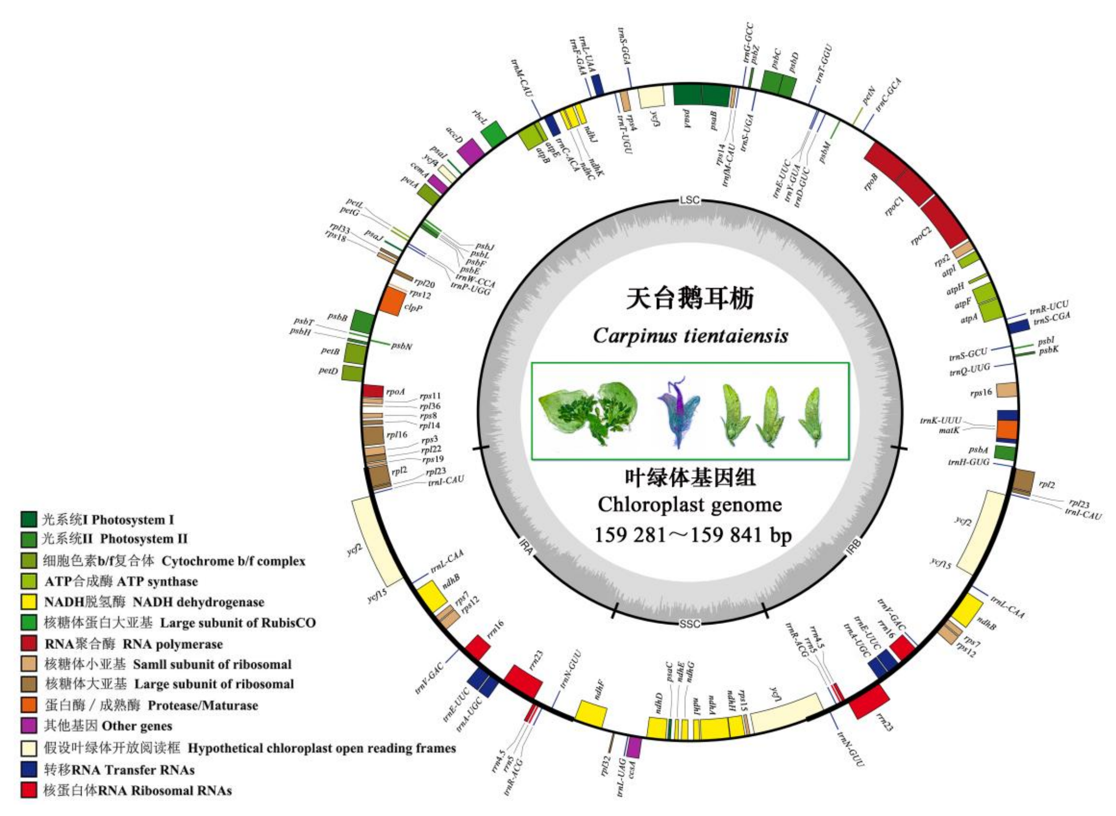
Illumina novaseq 6000 测序平台，以 150 bp 的对端读数，每个样本平均产生 3 000 万对配对读取，从 26 个天台鹅耳枥生成的数据中确定每个完整的 cp 基因组。天台鹅耳枥完整叶绿体基因组由四部分组成，与鹅耳枥属（*Carpinus*）其他植物相似（杨霄月，2019；赵儒楠等，2021）。长度为 159 281～159 841 bp，平均为 159 616.2 bp，每个叶绿体基因组由一个大的单拷贝区域（LSC）（88 360～88 711 bp，平均为 88 522.35 bp）和一个小的单拷贝区域（SSC）（18 420～18 794 bp，平均为 18 634.92 bp）组成，并由一对反向重复序列隔开（每个 IR 26 067～26 451 bp，平均为 26 229.46 bp）组成（表 2，图 1）。

表 2 天台鹅耳枥叶绿体基因组特征

Table 2 Characteristics of complete chloroplast genomes of *Carpinus tientaiensis*

基因组特征 Genome characteristic		数量 Amount
基因组大小	Genome size (bp)	159 616.2
大单拷贝区长度	LSC length (bp)	88 522.35
小单拷贝区长度	SSC length (bp)	18 634.92
反向重复序列长度	IR length (bp)	52 458.92
蛋白质编码基因数量	Protein-coding genes number	86
tRNA 基因数量	tRNA genes number	37
rRNA 基因数量	rRNA genes number	8
基因数	Genes number	131
GC 含量	GC content (%)	36.41

经过数据过滤后，Q20 质量评分（96.46％～97.21％）平均 96.98％，Q30 质量评分（90.81％～92.46％）平均 91.94％。基于叶绿体的统计，总体 GC 含量平均为 36.41％，LSC、SSC、IR 对应值分别为 34.20%、30.09%、42.37%，IR 区域的 GC 含量高于 LSC、SSC 区域。叶绿体基因组的 131 个基因中，包含 86 个蛋白编码基因、37 个 tRNA 基因和 8 个 rRNA 基因。其中：LSC 区域包含 60 个蛋白质编码和 22 个 tRNA 基因； SSC 区域包含 12 个蛋白质编码和 1 个 tRNA 基因；IR 区域有 7 个蛋白质编码，7 个 tRNA 和所有 4 个 rRNA 基因重复（表 2，图 1）。



不同功能群的基因用颜色编码。外圆的基因是顺时针转录的，而内圆的基因则是逆时针转录的。内圆的虚线区域表示绿体基因组的 GC 含量。

Genes belonging to different functional groups are color-coded. Genes shown outside the outer circle are transcribed clockwise and those inside are transcribed counterclockwise. Dashed area in the inner circle indicates the GC content of the chloroplast genome.

图 1 天台鹅耳枥叶绿体基因图谱

Fig. 1 Gene map of *Carpinus tientaiensis* chloroplast genome

2.3 叶绿体基因组的重复序列分析

滑动链的不匹配和重复序列的不适当重组可能导致序列变异和 DNA 重排（Wicke et al., 2011）。重复序列是重要的遗传标记，和物种的起源进化息息相关。重复序列一般可分为散在重复和简单序列重复。

散在重复使用 Reputer 软件分析，这里我们用 cpLTR 长末端重复序列表示。在天台鹅耳枥 cp 基因组中检测到正向重复平均 32 个，回文重复 25 个，反转重复 22 个（图 2），这些重复中的大多数表现出 10 到 38 bp 之间的长度。简单序列重复（SSR）广泛分布于叶绿体基因组中，包含长度为 1～6 bp 重复序列的短串联重复，叶绿体基因组中的这种短串联重复（SSR）是从单亲遗传过来的，SSR 在基因组重组和重排中发挥重要作用，在群体遗传和进化研究中，常被用作有效的分子标记（Zhou et al., 2018）。基于 SSR 位于叶绿体区段类型的数量统计，天台鹅耳枥蛋白编码基因平均为 48 个基因、tRNA 为 3 个基因、非编码区为 41 个基因。其中蛋白编码基因中 6 个基因（*matK*、*atpA*、*rpoB*、*atpB*、*cemA*、*rpl*2）具有 1 个串联重复，1 个基因 *rpoC*2 具有 4 个串联重复；1 个 *ycf*1 基因具有 2 个串联重复。

其中 LSC 和 SSC 区域共有 25 个 *rps*12 基因串联重复，IR 有 10 个 *rps*12 串联重复基因。

在 cp 基因组中有 87 个不同的 SSR 类型重复了 10 次以上，其中单核苷酸的数量最多，平均有 50 个，占总数的 57.47%；其次是四核苷酸为 14 个，二核苷酸为 12 个，三核苷酸、复合核苷酸各为 4 个，五核苷酸发现的比较少为 3 个，而六核苷酸没有发现（图 3）。在 cp 基因组中发现的 SSR 通常由 A / T 重复组成，很少包含 G / C 串联重复序列，其中单核苷酸由 A / T 碱基组成占 92%，这些 SSR 丰富了 cp 基因组的 AT。cp SSR 碱基重复序列中 10 条占 31.20%，11 条占 16.95%、12 条占 34.44%、13 条占 5.65%、14 条以上占 11.76%（图 4）。cp SSR 分区 IRA、IRB 各占 6.58%，LSC 占 68.59%，SSC 占 17.96%，其中 LSC 重复序列最多（图 5）。在天台鹅耳枥 cp DNA 中，观察到 SSR 具有丰富的碱基重复和重复条数，可以作为居群进化研究有用的遗传信息。

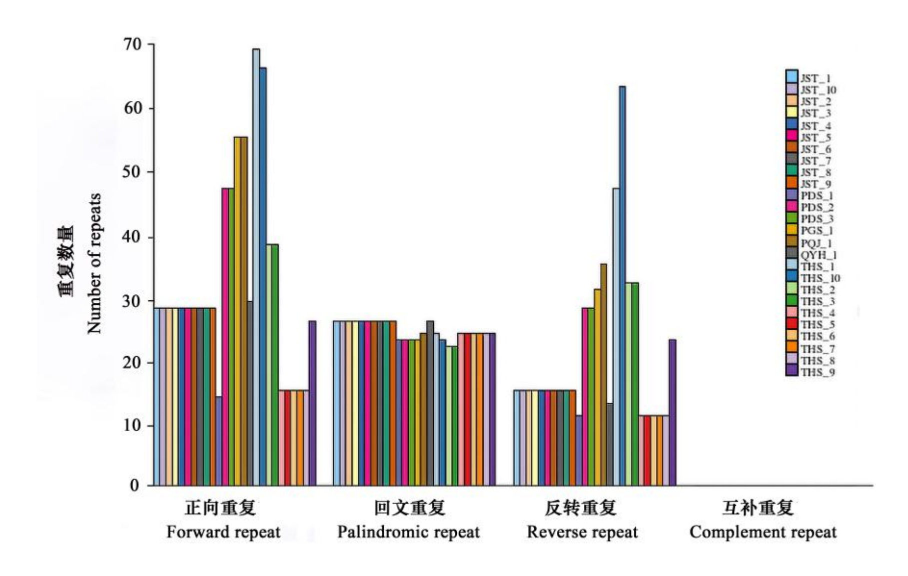
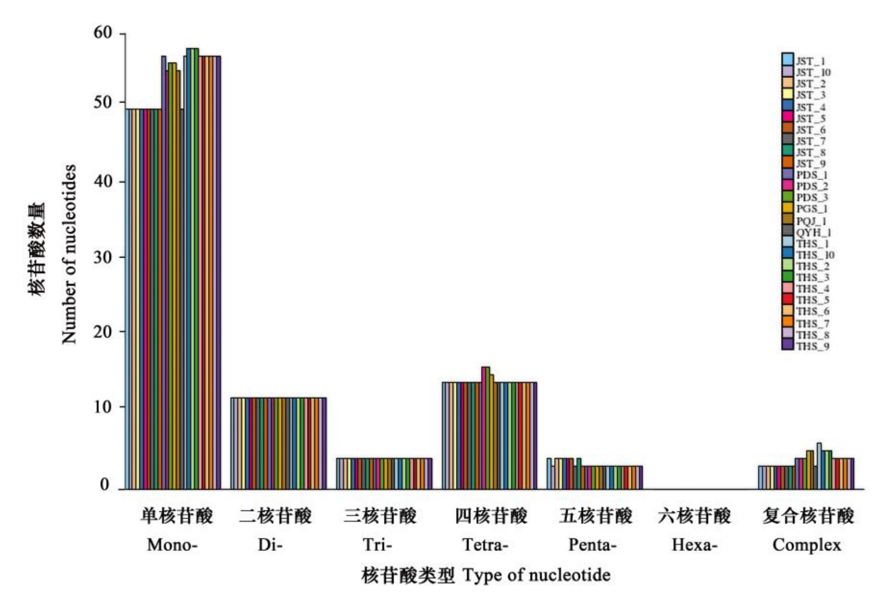


图 2 天台鹅耳枥 cp LTR 的类型数量

Fig. 2 Number of cp LTR types detected in chloroplast genomes of *Carpinus tientaiensis*



Mono-, Di-, Tri-, Tetra-, Penta-, Hexa-表示 1～6 核苷酸 SSR，Complex 为复合核苷酸 SSR。

Mono-, Di-, Tri-, Tetra-, Penta-, Hexa- indicate 1 to 6 nucleotide SSR, Complex indicates nucleotide complex SSR.

图 3 天台鹅耳枥 cp SSR 的类型数量

Fig. 3 Number of cp SSR types detected in chloroplast genomes of *Carpinus tientaiensis*

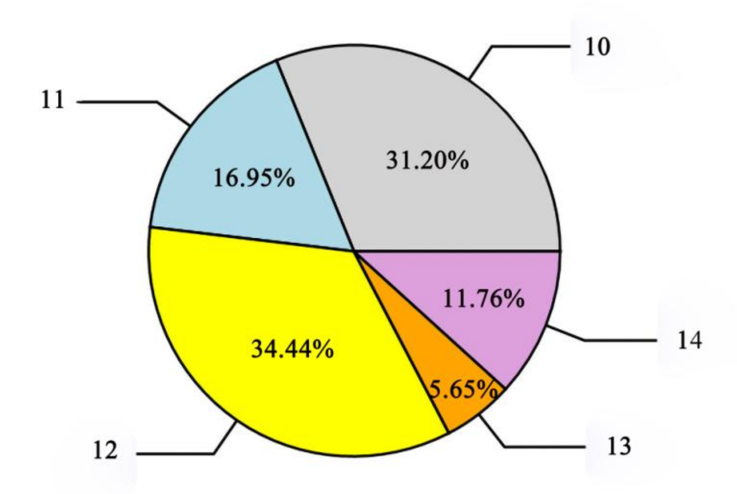


图 4 天台鹅耳枥 cp SSR 碱基条数分布

Fig. 4 Frequencies of identified of *Carpinus tientaiensis*

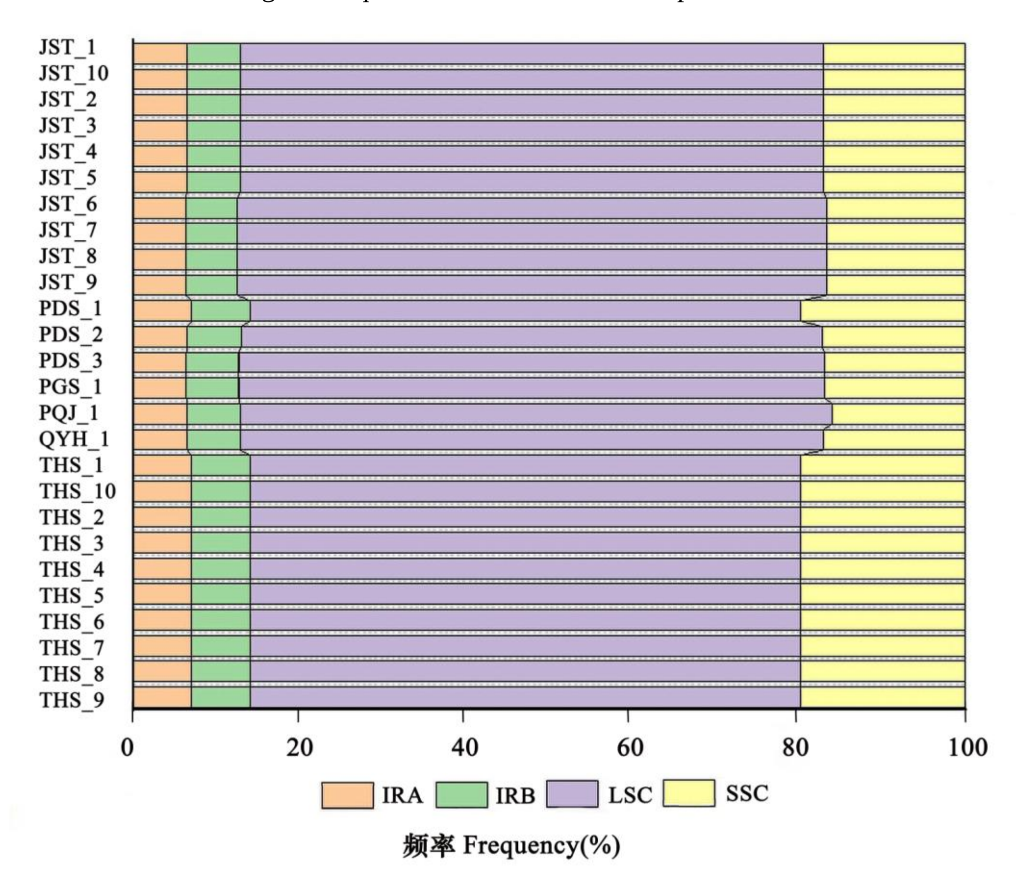


图 5 天台鹅耳枥 SSR 在 LSC、IR 和 SSC 区域的频率

Fig. 5 Frequencies of identified SSR in LSC, IR and SSC regions of *Carpinus tientaiensis*

2.4 密码子偏好性分析

密码子在遗传信息的传递中起着重要作用，是核酸和蛋白质之间的联系。通过对所有蛋白质编码的叶绿体 DNA 和氨基酸序列进行统计分析，天台鹅耳枥叶绿体基因组序列的 65.65％是蛋白编码基因，结果显示蛋白质密码子的相似性，其中 AUG、UUA、AGA、GCU、UCU 的频率较高，而 CUG、GUG、AGC、CUC、CUG 的频率较低（图 6）。在这些密码子中，蛋白质编码基因中最常见的氨基酸是 Ile（异亮氨酸），其在叶绿体 DNA 中出现 1 146 次。相对同义密码子使用（RSCU）值分析表明，Trp（色氨酸）蛋白质编码基因 RSCU = 1，表示该密码子没有偏好性；其中 47.62%密码子的 RSCU> 1，大多数（28/30，93.33％）以 A 或 T（U）结尾；50.79%密码子的 RSCU <1，大多数（30/32，93.75％）以 C 或 G 结尾（图 6）。

在天台鹅耳枥物种的蛋白质编码叶绿体基因中，20 个氨基酸由 63 个密码子编码，其中除 Asp（天冬氨酸）外，大多数氨基酸都具有密码子偏好性。总共确定了 40 个密码子偏好，其中涉及 19 个氨基酸。在优选的密码子中，63.49％表现出较高的偏好（图 7）。该结果进一步揭示了天台鹅耳枥叶绿体基因组的相对保守性，因为高密码子偏好也是高等植物中的常见现象（喻凤等，2021）。

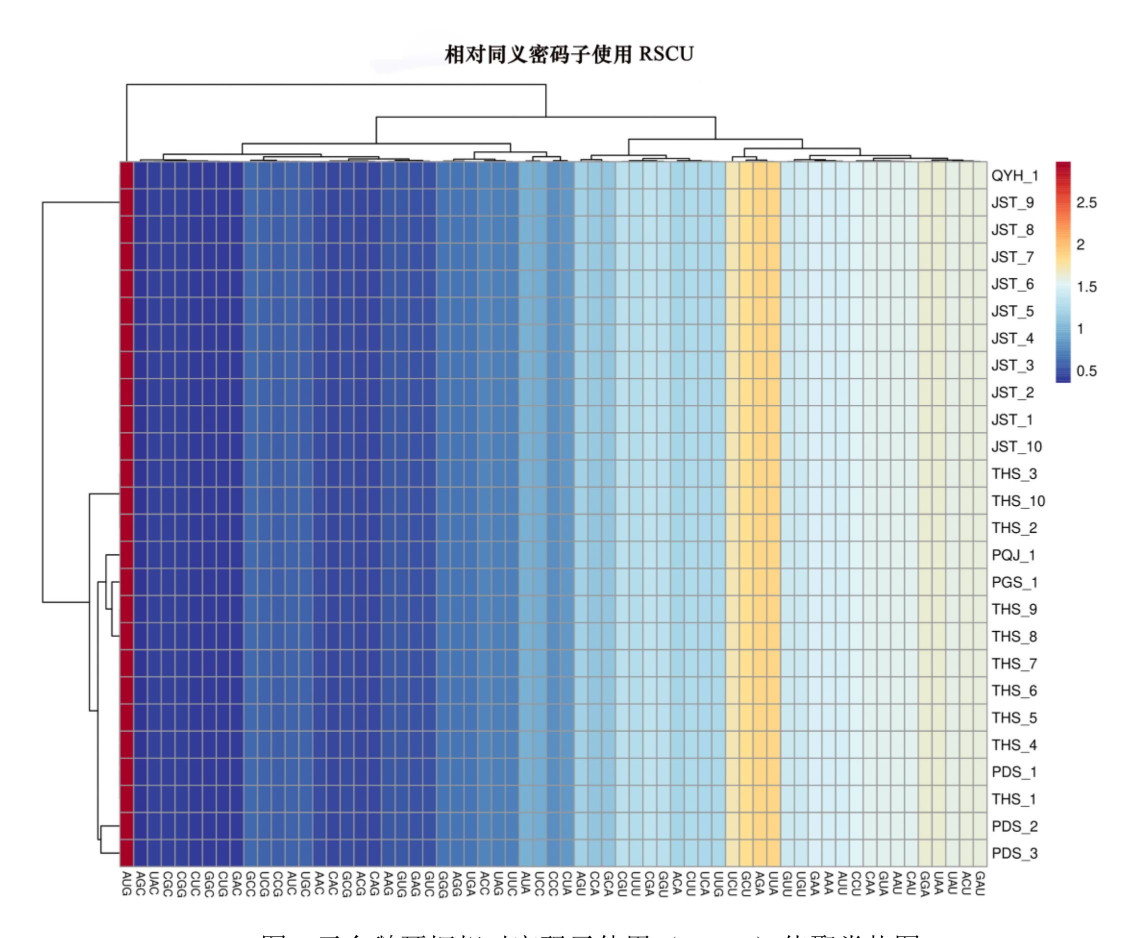
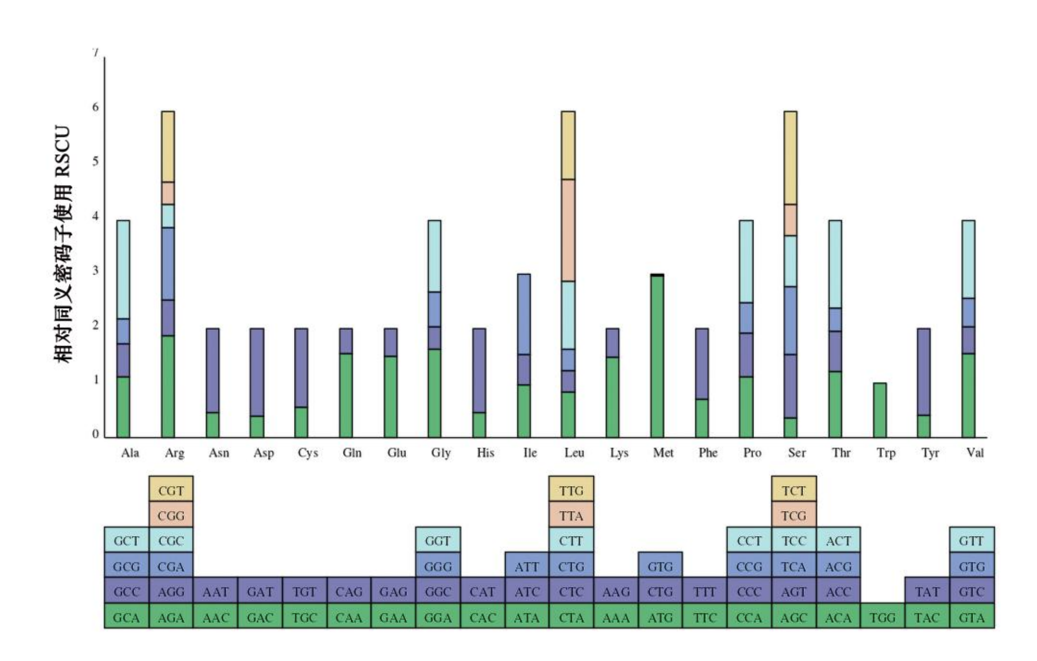


图 6 天台鹅耳枥相对密码子使用（RSCU）值聚类热图

Fig.6 Heat map of the relative synonymous codon usage (RSCU) value of *Carpinus tientaiensis*



Ala. 丙氨酸；**Arg**. 精氨酸；**Asn**. 天冬酰胺；**Asp**. 天冬氨酸；**Cys**. 半胱氨酸；**Gln**. 谷氨酰胺；**Glu**. 谷氨酸；**Gly**. 甘氨酸；**His**. 组氨酸；**Ile**. 异亮氨酸；**Leu**. 亮氨酸；**Lys**. 赖氨酸；**Met**. 甲硫氨酸；**Phe**. 苯丙氨酸；**Pro**. 脯氨酸；**Ser**. 丝氨酸；**Thr**. 苏氨酸；**Trp**. 色氨酸；**Tyr**. 酪氨酸；**Val**. 缬氨酸。

Ala. Alanine; **Arg**. Argnine; **Asn**. Asparagine; **Asp**. Aspartate; **Cys**. Cysteine; **Gln**. Glutamine; **Glu**. Glutamate; **Gly**. Glycine; **His**. Histidine; **Ile**. Isoleucine; **Leu**. Leucine; **Lys**. Lysine; **Met**. Methionine; **Phe**. Phenylalanine; **Pro**. Proline; **Ser**. Serine; **Thr**. Threonine; **Trp**. Tryptophan; **Tyr**. Tyrosine; **Val**. Valine.

图 7 天台鹅耳枥蛋白质编码基因中 20 个氨基酸密码子含量

Fig. 7 Codon content of 20 amino acid in all protein-coding genes of *Carpinus tientaiensis*

2.5 叶绿体基因组变异检测

单核苷酸多态性（SNP）和基因组结构变异在进化过程中至关重要（Britten et al., 2003）。叶绿体基因组结构性变异有插入 / 删除、转换、颠换和基因组结构重排。在 26 个天台鹅耳枥 cp DNA 中鉴定了 314 个单核苷酸多态性（SNP），SNP 的数量 JST 居群平均为 132 条、QYH 为 126 条、PDS 为 18 条、PGS 为 18 条、PQJ 为 8 条、THS 为 12 条。以居群植株与 THS _1 叶绿体基因组比对，结果表明碱基替换数（Tv）JST 居群平均为 95 个、QYH 为 90 个、PDS 为 16 个、PGS 为 15 个、PQJ 为 6 个、THS 为 12 个（表 3）。研究的所有成对序列比较表明，颠换次数多于转换次数，这在其他分类群中也被发现（Stoltzfus et al., 2016）。

表 3 天台鹅耳枥叶绿体比较基因组统计表

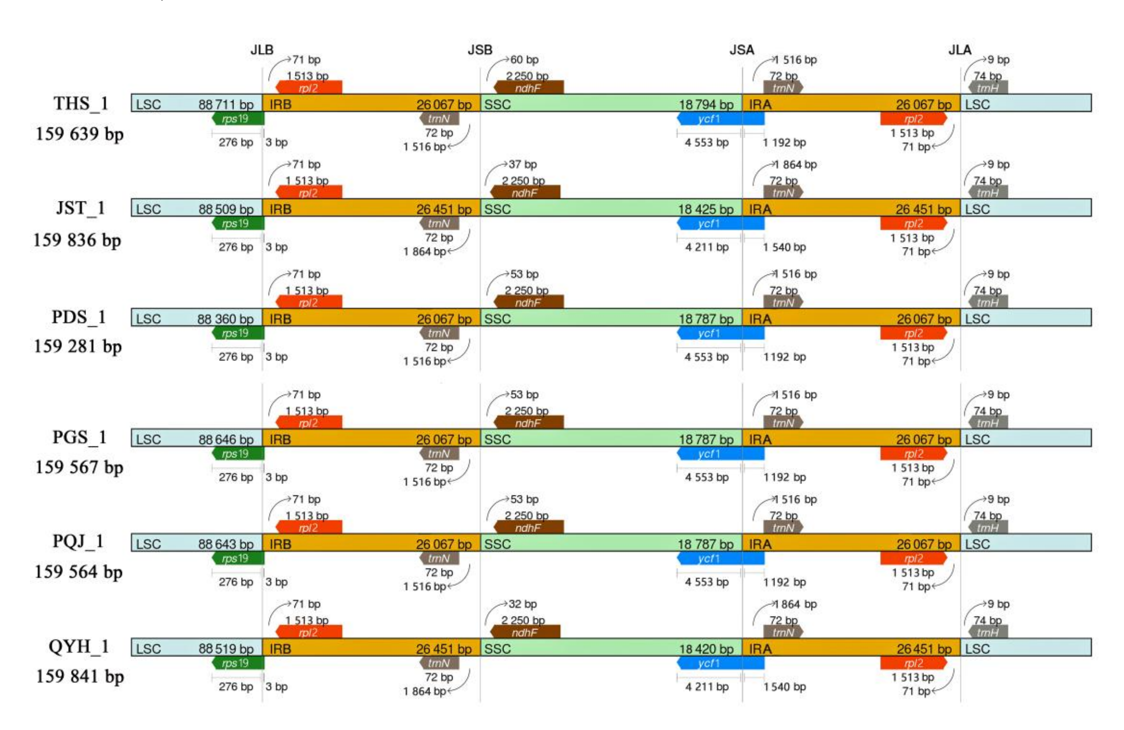
Table 3 Chloroplast comparative genome statistical table of *Carpinus tientaiensis*

居群 Population	单核苷酸多态性 Single nucleotide polymorphism (SNPs)	插入和缺失 Insertion and deletion (INDEL)	转换 Transition (Ts)	颠换 Transversion (Tv)	插入 Insertion (INS)	缺失 Deletion (DEL)
JST	132	90	37	95	47	43

QYH	126	99	36	90	39	60
PDS	18	20	3	16	5	15
PGS	18	18	3	15	4	14
PQJ	8	11	2	6	1	10
THS	12	17	0	12	5	12
合计 Total	314	255	81	234	101	154

2.6 IR 区域与边界的扩张收缩

被子植物的 cp 基因组是高度保守的，IR 区域和单拷贝（SC）边界区域的扩展和收缩是造成高等植物 cp 基因组长度变化的主要机制（Saina et al., 2018）。比较 26 个完整的天台鹅耳枥 cp 基因组的 IR / SC 边界区域，发现在连接位置上有明显差异。其中 *rpl*2 基因的长度都是 1 513 bp，其端点较保守，离 JLA、JLB 基因间隔区都是 71 bp。*trnN-GUU* 的长度是 72 bp，离 JSB、JSA 基因间隔区变化分别为 1 516～1 864 bp，不超过以下基因间隔区的终点。*rps*19 基因在 JSB / IRb 边界处由 JSB 向 IRb 延伸 3 bp。*ycf*1 基因跨越 JSA / IRa 区，4 211～4 553 bp 位于 SSC 区域内，向 IRb 延伸 1 192～1 540 bp（图 8）。这四个 cp 基因组中 IR / SC 边界处的变异导致了整个 cp 基因组序列长度的差异，并在其他植物中也已发现（Yin et al., 2018）。



基因用彩色框表示。基因特征上方的数字表示基因末端与边界位点之间的距离。

Genes are denoted by colored boxes. The numbers above the gene features indicate the distance between the ends of genes and the border sites.

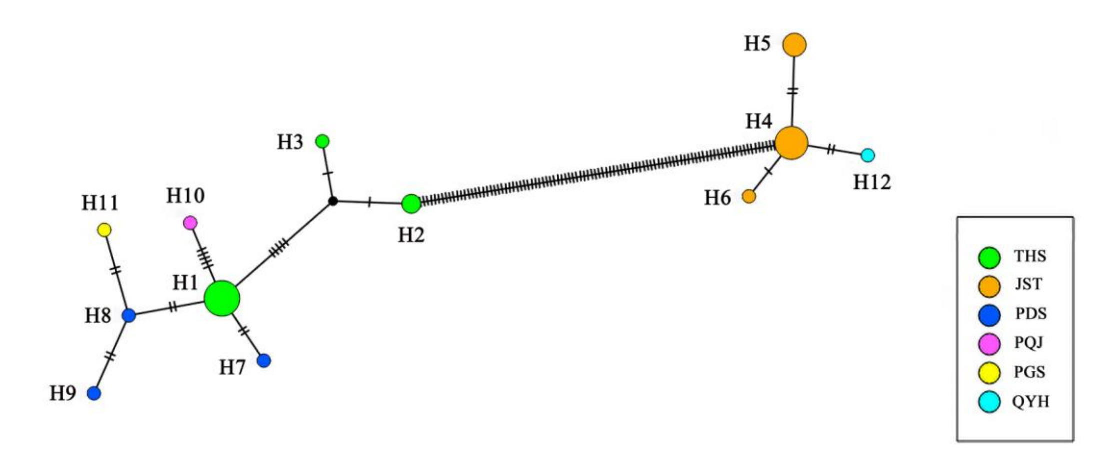
图 8 天台鹅耳枥叶绿体基因组的 IR-SC 区的边界比较

Fig. 8 Comparison of IR-SC border positions across chloroplast genomes of *Carpinus tientaiensis*

2.7 基于完整的叶绿体基因组序列的群体分析

2.7.1 单倍型网络构建

通过天台鹅耳枥 cpDNA 分析，运用一段遗传连锁的核酸序列（SNP）的变异来区分，构建单倍型网络（haplotype network）。在系统发育分析时，核苷酸碱基一个插入或缺失作为一次进化事件进行编码分析。天台鹅耳枥单倍型中的大多数具有居群单倍型，几乎没有观察到居群之间的单倍型共享。在地理上接近的天台县（THS）H1～H3、景宁县（JST）H4～H6、磐安县大盘山（PDS）H7～H9，各居群内具有相同单倍型；而磐安县青明尖（PQJ）H10、磐安县高姥山（PGS）H11 和青田县（PQJ）H12 是在单个标本中发现的私有单倍型。根据统计单倍型网络图（图 9），揭示了两组单倍型，在此研究中称为浙东组 THS 居群和浙西组 JST 居群，由 95 个突变步骤隔开。根据多态位点分析 JST 居群和 THS 居群可变位点和插入，在 *rps*16-*trn*Q 序列有三个核苷酸取代分别为 T→C（位置 275）、A→G（位置 299）和 T→G（位置 1254）区分了浙东组 THS 居群（T、A 和 T）和浙西组 JST 居群（C、G 和 G）。从网络图推测可能的种群历史，除来自 THS 居群和 JST 居群单倍型之间碱基突变为 95 个较远外，居群组单倍型之间演化关系呈现星状中心辐射。THS 居群的单倍型演化出 PDS 亚群、PQJ 亚群和 PGS 亚群，居群组单倍型之间差异仅 2～6 个碱基突变，THS 居群与 PDS 亚群单倍型之间的替代链接提示可能存在同源性；JST 居群的单倍型演化出 QYH 亚型，2 个单倍型之间少量差异（1～2 个碱基突变）；显示出天台鹅耳枥居群在历史上遇到瓶颈后曾发生局部扩张。



网络图中每个圆圈代表一个单倍型，每个圆圈的大小与观察到的频率成比例，圆圈颜色代表不同的居群，分支上的短线表示单倍型间的碱基替换数。

Each circle in the network represents a haplotype, the size of the circle represents the frequency of haplotypes, and different colors in the circle represent different populations. and short lines on branches represent base substitutions between haplotypes.

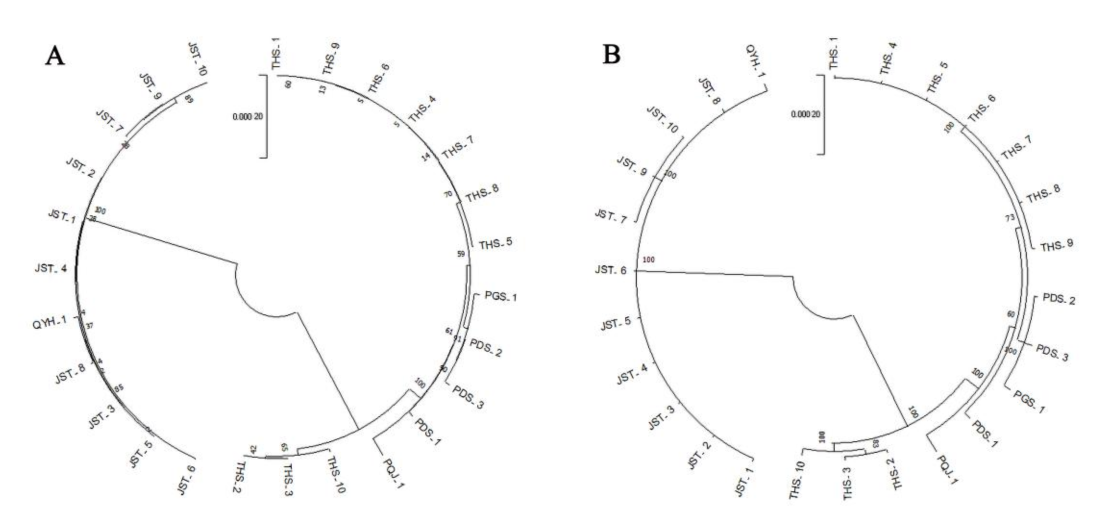
图 9 基于 cpDNA 的天台鹅耳枥单倍型网络

Fig. 9 Network of haplotypes of *Carpinus tientaiensis* based on cpDNA

2.7.2 系统进化树构建

叶绿体基因组包含丰富的系统发育信息，已被广泛用于近缘种间和种内水平的系统发育重建。使用叶绿体基因组数据，解决了与各种系统发育困难群体相关的长期争议。为了评估天台鹅耳枥系统发生关系，使用 26 个天台鹅耳枥植株全叶绿体序列进行系统发育分析。通过极大似然法（RAxML）和贝叶斯推断法（MrBayes）生成系统树，系统发育分析每个节点数字为极大似然法（ML）引导值（BS）（%）和贝叶斯（BA）后验概率（PP）（%）（Stamatakis, 2014；Xie et al., 2018）。

在完整叶绿体基因组序列中，通过系统分析，6 个自然居群分为 THS 居群和 JST 居群（100BS/100PP）。THS 居群包含第一分支 THS_10 和 THS_2～3（65BS/83PP）；第二分支 PQJ_1（61BS/60PP），PDS_1（59BS/73PP），PGS_1 和 PDS_2～3（70BS/100PP），THS_1 和 THS_4～9（70BS/100PP）。JST 居群第一分支 JST_7 和 JST_9～10（89BS/100PP）；第二分支 JST_1～6、JST_8 和 QYH_1（85BS/100PP）（图 10）。LSC、SSC 数据集的拓扑结构与物种叶绿体基因组系统树具有一致性，在种内进化枝中仅发生了细微的拓扑差异，支持形成单系的群体。



A. cpDNA ML 系统发育树；**B**. cpDNA Bayes 系统发育树。分支数表示 ML 支持值 / Bayes 后验概率。

A. cpDNA ML tree system development tree；**B**. cpDNA Bayes tree system development tree. Number of the branches indicate ML bootstrap support value/Bayesian posterior probability.

图 10 基于完整叶绿体基因组序列的最大似然和贝叶斯推理方法重建 6 个分类群的系统发育树

Fig. 10 Phylogenetic tree reconstruction of six taxa using maximum likelihood and Bayesian inference methods based on the complete chloroplast genome sequences

2.7.3 谱系结构分析

利用 DnaSp 计算遗传多样性参数，通过 AMOVA 分子方差分析，可以评估谱系间的分化变异程度，遗传分化指数 F_{st} 分析可以得到谱系间的分化程度。核苷酸取代和插入/缺失变异揭示了 6 个天台鹅耳枥分布区中的 12 个 cpDNA 单倍型（N_h），这些单倍型中的 25%是在单个植株中发现的，几乎没有观察到居群之间的共享单倍型。谱系间的分化可以通过

对单倍型多样性指数（H_d）与核苷酸多样性指数（P_i）进行分析，其数值越大，说明其遗传多样性越高（周文漪，2014；Nikulin et al., 2020）。其中 PGS 亚群、PQJ 亚群和 YHS 亚群由单一植株组成，仅有私有单倍型。居群核苷酸多样性指数（P_i），JST 居群 P_i 为 0.000 01，THS 居群 P_i 为 0.000 02，PDS 居群 P_i 为 0.000 03，所有居群核苷酸多样性的变异均较低（P_i<0.005）。从单倍型多样性指数看，JST 居群 H_d 为 0.6，THS 居群 H_d 为 0.511， JST 居群和 THS 居群单倍型多样性相对较低，其中 PDS 亚群 H_d 为 1，由于 PDS 居群仅有 3 株个体，单倍型多样性偏高（表 4）。

表 4 谱系多样性参数统计表

Table 4 Statistical table of genealogical diversity parameters

居群 Population	样本数 Number of samples (N)	分比单倍型个数 Number of haplotypes (N_h)	分离位点数目 Polymorphic sites (S)	单倍型多样性 Haplotype diversity (H_d)	平均核苷酸差异 Average number of nucleotide differences (K)	核苷酸多样性 Nucleotide diversity (P_i)
JST	10	3	3	0.6	1.133	0.000 01
PDS	3	3	6	1	4	0.000 03
THS	10	3	9	0.511	3.822	0.000 02

天台鹅耳枥 6 个居群的分子变异分析（AMOVA）结果表明，居群间的遗传分化固定指数 F_{st} 为 0.970 9，谱系间分化较大，表明在整个遗传变异中居群间遗传变异占 97.09%，居群内遗传变异占 2.91%，居群间的遗传变异大于居群内（表 5）。这种分布格局的主要原因是生境的片段化，地理隔离阻碍了居群间的基因交流（孙逸，2012；郑鑫，2015；Nikulin et al., 2020）。

表 5 谱系分子方差检验统计表

Table 5 Statistical table of Genealogical molecular variance test

变异来源 Source of variation	自由度 df	平方和 Sum of squares	平方差 Variance squares	百分比 Percentage（%）
居群间 Among populations	5	662.4	36.807 65	97.09
居群内 Within populations	20	22.1	1.105	2.91
总计 Total	25	684.5	37.912 65	
固定指数 Fixation index (F_{st})			0.970 85	

3 讨论与结论

完整的叶绿体基因组序列可提供丰富的系统发育信息来源，天台鹅耳枥 cpDNA 的长度平均为 159 616.2 bp，GC 含量相似为 36.41%，说明该物种 cp 基因组的高度保守性质。重复分析显示在天台鹅耳枥 cp DNA 中发现 LTR 重复序列包括正向重复平均为 32 个，回文重复为 25 个，反转重复为 22 个；SSR 重复序列不同类型为 87 个。这些重复中的大多数位于蛋白编码区、非编码区和 tRNA 中。在藻类和被子植物基因组中 LTR 重复序列很常见，

是促进 cp 基因组重排的主要因素之一，并且许多重排终点都与此类重复序列相关（Pombert et al., 2005；Zhang et al., 2020）。在所有个体中，SSR 通常由 A/T 重复组成，其中单核苷酸由 A/T 碱基组成占 92%，大多数蛋白质编码基因都具有高度的密码子偏好性，在优选的密码子中，63.49%表现出较高的偏好，叶绿体密码子的第三个 A/T 偏好较高。相关研究表明基因组 AT 含量与重复序列的动力学以及叶绿体蛋白编码基因的密码子偏向性有关（Yu et al., 2019；Wu et al., 2020）。

叶绿体基因组 SNP 的数量及碱基替换为物种之间的系统发育解析提供了有益的标记（郑鑫，2015；Nikulin et al., 2020）。根据溯祖理论，通过单倍型网络图，结合地理信息我们可以推断种群的起源、扩散历史（Huang et al., 2014）。通过对 SNP 多态位点和核苷酸变异分析，单核苷酸取代显示天台鹅耳枥分为天台县居群（THS）和景宁县居群（JST），亚群间距离相近的优先聚为一支。除来自 THS 居群和 JST 居群单倍型关系较远外，居群组大部分单倍型之间仅有 2～6 个碱基突变，某几个同一或者相近地理居群的单倍型具有聚类现象，这可能由于该物种对亚热带湿润气候或暖温带气候的环境要求较高造成（陈之端，1994）。天台鹅耳枥具有居群组单倍型，没有居群之间的单倍型共享，这种单系进化枝可能是由于地理隔离导致短时期内居群之间基因交流较少造成的（Nikulin et al., 2020）。所有居群核苷酸多样性的变异均较低（P_i＜0.005），JST 居群和 THS 居群单倍型多样性较低（H_d 为 0.5～0.6），意味着最近发生过居群的瓶颈效应（孙逸，2012；周文漪，2014）。由于晚第三纪及第四纪的气候动荡及冰期更替，天台鹅耳枥退缩到狭小区域的避难所，度过冰期；或群体局部扩张到湿润地区的森林中（陈之端，1994；Qi et al., 2012）。

通过对天台鹅耳枥谱系结构与分化的研究，对遗传多样性比较高的居群制定资源保存和引种驯化策略。在 6 个自然居群中都检测到了独特的单倍型，尤其 THS 居群、JST 居群和 PDS 亚群具有较高的单倍型多样性，栖息地面积较大内部环境较稳定，需要加强栖息地保护，以维持天台鹅耳枥高的遗传多样性。天台鹅耳枥居群规模较小、隔离程度较高，居群间呈现较大的遗传分化，致使种源不断减少，是急需保护的濒危植物。对繁殖衰退的居群应开展遗传拯救，引入以花粉为主导基因流实验，移入新个体或基因型而减缓遗传侵蚀进而提高天台鹅耳枥种群生存力。

参考文献：

AGUILAR R, QUESADA M, ASHWORTH L, et al., 2008. Genetic consequences of habitat fragmentation in plant populations: susceptible signals in plant traits and methodological approaches[J]. Mol Ecol, 17(24): 5177-5188.

BIRKY CW, 1995. Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution[J]. Proc Natl Acad Sci USA, 92(25): 11331-11338.

BRITTEN RJ, ROWEN L, WILLIAMS J, et al., 2003. Majority of divergence between closely related DNA samples is due to indels[J]. Proc Natl Acad Sci USA, 100(8): 4661-4665.

CHEN MS, KE SS, 2013. Acclimation of anatomical structure and photosynthesis characteristics in leaves of *Carpinus tientaiensis* to irradiance[J]. Sci Silv Sin, 49(2): 46-53. [陈模舜，柯世省，2013. 天台鹅耳枥叶片的解剖结构和光合特性对光照的适应[J]. 林业科学，49(2): 46-53.]

CHEN MS, KE SS, JIN ZX, et al., 2020. Conservation biology of *Carpinus tientaiensis*[M]. Beijing: China Forestry Publishing House: 196-212. [陈模舜，柯世省，金则新，等，2020. 天台鹅耳枥保护生物学[M]. 北京：中国林业出版社：196-212.]

CHEN ZD, 1994. Phylogeny and phytogeography of the Betulaceae(II) [J]. Acta Phytotax Sin, 32(2): 101-153. [陈之端，1994. 桦木科植物的系统发育和地理分布(续) [J]. 植物分类学报，32(2): 101-153.]

FU LK, 2003. Higher plants of china: Vol. 4[M]. Qingdao: Qingdao Publishing House: 260-270.[傅立国，2003. 中国高等植物: 第四卷[M]. 青岛：青岛出版社：260-270.]

HUANG DQ, LI QQ, ZHOU CJ, et al., 2014. Intraspecific differentiation of *Allium wallichii* (Amaryllidaceae) inferred from chloroplast DNA and internal transcribed spacer fragments[J]. J Syst Evol, 52(3): 341-354.

KIMURA M, 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences[J]. J Mol Evol, 16(2):111-120.

LEIGH JW, BRYANT D, 2015. PopART: Full-feature software for haplotype network construction[J]. Methods Ecol Evol, 6(9):1110-1116.

NIKULIN AY, NIKULIN VY, GONTCHAROV AA, 2020. *Orostachys spinosa*(Crassulaceae) origin and diversification: East Asia or South Siberian Mountains? Chloroplast DNA data[J]. Plant Syst Evol, 306(5): 84.

POMBERT JF, OTIS C, LEMIEUX C, et al., 2005. The chloroplast genome sequence of the green alga *Pseudendoclonium akinetum*(Ulvophyceae) reveals unusual structural features and new insights into the branching order of chlorophyte lineages[J]. Mol Biol Evol, 22(9): 1903-18.

QI XS, CHEN C, COMES HP, et al., 2012. Molecular data and ecological niche modelling reveal a highly dynamic evolutionary history of the East Asian Tertiary relict *Cercidiphyllum* (Cercidiphyllaceae)[J]. New Phytol, 196(2): 617-630.

SAINA JK, LI ZZ, GICHIRA AW, et al., 2018. The complete chloroplast genome sequence of tree of heaven *Ailanthus altissima* (Mill.) (Sapindales: Simaroubaceae), an important pantropical tree[J]. Int J Mol Sci, 19(4): 929.

SOLTIS PM, SOLTIS DE, 2000. The role of genetic and genomic attributes in the success of polyploids[J]. Proc Natl Acad Sci USA, 97(13): 7051-7057.

STAMATAKIS A, 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies[J]. Bioinformatics, 30(9): 1312-1313.

STOLTZFUS A, NORRIS RW, 2016. On the causes of evolutionary transition: transversion bias[J]. Mol Biol Evol, 33(3):595-602.

SUN Y, 2012. Phylogeography and population genetics of the East Asia endemic genus *Kirengeshoma* (Hydrangeaceae)[D]. Hangzhou: Zhejiang University. [孙逸，2012. 东亚特有濒危植物黄山梅的亲缘地理学与群体遗传学研究[D]. 杭州：浙江大学.]

WANG CT, YE CL, 2007. Endangering causes of endemic rare wild plants and conservation methods in Zhejiang Province[J]. J Fujian For Sci Technol, 34(2): 202-204.[王昌腾，叶春林, 2007. 浙江省特有野生珍贵植物濒危原因及保护对策[J]. 福建林业科技，34(2)：202-204.]

WANG JH, 2015. Preliminary study on growth performance and molecular phylogeography of natural populations in an endangered maple, *Acer griseum* (Dicotyledoneae: Sapindaceae), endemic to China[D]. Beijing: Chinese Academy of Forestry Sciences. [王佳慧，2015. 特有濒危种血皮槭天然群体生长状况及分子谱系地理初步研究[D]. 北京：中国林业科学研究院.]

WEI XZ J, IANG MX, 2012. Limited genetic impacts of habitat fragmentation in an “old rare” relict tree, *Euptelea pleiospermum* (Eupteleaceae)[J]. Plant Ecol, 213(6): 909-917.

WICKE S, SCHNEEWEISS GM, DEPAMPHILIS CW, et al., 2011. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function[J]. Plant Mol Biol, 76(3-5): 273-297.

WU ZH, LIAO R, YANG TG, et al., 2020. Analysis of six chloroplast genomes provides insight into the evolution of *Chrysosplenium* (Saxifragaceae)[J]. BMC Genomics, 21(1): 621.

WYMAN SK, JANSEN RK, BOORE JL, 2004. Automatic annotation of organellar genomes with DOGMA[J]. Bioinformatics, 20(17): 3252-3255.

XIE DF, YU Y, DENG YQ, et al., 2018. Comparative analysis of the chloroplast genomes of the Chinese endemic genus *Urophysa* and their contribution to chloroplast phylogeny and adaptive evolution[J]. Int J Mol Sci, 19(7):1847.

YANG XY, 2019. Phylogenetic analysis of betulaceae plastomes[D]. Lanzhou: Lanzhou University: 47-52. [杨霄月，2019. 桦木科叶绿体基因组的系统发育分析[D]. 兰州：兰州大学：47-52.]

YANG YZ, WANG MC, LU ZQ, et al., 2017. Characterization of the complete chloroplast genome of *Carpinus tientaiensis*[J]. Conserv Genet Resour, 9(2): 339-341.

YIN KQ, ZHANG Y, LI YJ, et al., 2018. Different natural selection pressures on the *atpF* gene in evergreen sclerophyllous and deciduous oak species: evidence from comparative analysis of the complete chloroplast genome of *Quercus aquifolioides* with other oak species[J]. Int J Mol Sci, 19(4): 1042.

YU F, HAN M, 2021. Analysis of codon usage bias in the chloroplast genome of alfalfa (*Medicago sativa*)[J]. Guihaia, 41(12): 2069-2076. [喻凤，韩明，2021. 紫花苜蓿叶绿体基因组密码子偏好性分析[J]. 广西植物，41(12): 2069-2076.]

YU T, ZHANG YY, GAO J, et al., 2019. Compkete chloroplast genome sequence of *Betula*

halophila, a plant species with estremely small populations[J]. Sci Silv Sin, 55(2): 41-49. [于涛，张宇阳，高健，等，2019. 极小种群濒危植物盐桦叶绿体基因组特征分析[J]. 林业科学，55(2): 41-49.]

ZHANG QJ, LI W, LI K, et al., 2020. The chromosome-level reference genome of tea tree unveils recent bursts of non-autonomous LTR retrotransposons in driving genome size evolution[J]. Mol Plant, 13(7): 935-938.

ZHANG SY, DING BY, 1993. Flora of Zhejiang (Volume General)[M]. Hangzhou: Zhejiang Science and Technology Publishing House: 245-251.[章绍尧，丁炳扬，1993. 浙江植物志(总论卷)[M]. 杭州： 浙江科学技术出版社： 245-251.]

ZHAO RN, CHU XJ, LIU W, et al., 2021. Structure and variation analysis of chloroplast genomes in *Carpinus* [J]. J Nanjing For Univ (Nat Sci Ed), 45(2): 25-34. [赵儒楠，褚晓洁，刘维，等，2021. 鹅耳枥属树种叶绿体基因组结构及变异分析[J]. 南京林业大学学报(自然科学版)，45(2)：25-34.]

ZHENG X, 2015. Origin area and glacial refugia: chloroplast DNA diversity in the arctic-alpine plant *Oxyria digyna* (Polygonaceae)[D]. Hefei: Anhui University. [郑鑫，2015. 基于分子信息推测北极高山植物山蓼的起源地及冰期避难所[D]. 合肥：安徽大学.]

Editorial Committee of the Flora of China of Chinese Academy of Science, 1979. Flora Reipublicae Popularis Sinicae (Vol.21)[M]. Beijing: Science Press: 65-67. [中国科学院中国植物志编辑委员会，1979. 中国植物志(第二十一卷) [M]. 北京：科学出版社：65-67.]

ZHOU T, WANG J, JIA Y, et al., 2018. Comparative chloroplast genome analyses of species in *Gentiana* section *Cruciata* (Gentianaceae) and the development of authentication markers[J]. Int J Mol Sci, 19(7): 1962.

ZHOU WY, 2014. MtDNA Cytb diversity of *Siniperca scherceri* from seven water systems[D]. Guangzhou: Ji’nan University：41-46. [周文漪，2014. 基于线粒体细胞色素 b 的 7 水系斑鳜遗传多样性分析[D]. 广州：暨南大学：41-46.]